CLAIMS

1. A computer natural language translation system, comprising:

means for inputting source language text;

means for outputting target language text; and

transfer means for generating said target language text from said source language text using stored translation data generated from examples of source and corresponding target language texts, the transfer means being arranged to use data defining a plurality of stored translation units each consisting of a small number of ordered words and/or variables in both the source and the target language, and

development means for inputting new examples of source and corresponding target language texts, and adding new translation units based thereon,

the development means being arranged:

to apply said stored translation data to a new example of source and corresponding target language texts, to generate for each at least one analysis comprising analysis data indicating the dependencies of words therein;

to calculate, for each one of a plurality of source words in the source language text, a measure of affinity between each word in the target language text and each such source language word;

to pair source language words with target language words on the basis of the measures thus calculated, and

to form new translation units comprising a said paired word and those words and/or variables in both the source and the target language analyses which depend upon it.

2. A system according to claim 1, in which the development means is arranged to be capable of generating a plurality of said analyses in at least one of the source and target language, and to select one pair of analyses from which to form said new translation units.

3. A system according to claim 2, in which the development means is arranged to jointly select the pair of analyses and the pairing of said source and target words.

4. A system according to any preceding claim, in which said analysis data represents, or can be converted into, a tree structure indicating the dependencies of words therein

5. A system according to any preceding claim, in which the development means is arranged to perform said analyses using the stored translation units.

6. A system according to any preceding claim, in which the development means is arranged to calculate said measures of affinity using the stored translation units.

7. A system according to any preceding claim, in which the development means is arranged to calculate said measures of affinity using a lexicon database through which translations in said source and target languages can be identified.

8. A system according to any preceding claim, in which the development means is arranged to calculate said measures of affinity using semantic and/or syntactic analysis.

9. A system according to any preceding claim, wherein the measure of affinity is a measure of the probability that each word in the target language text is a translation of each respective source language word.

10. A system according to any preceding claim, in which the development means is arranged to perform said pairing in order of probability of correspondence from the highest probability, using said measures of probability.

11. A system according to claim 10, in which, after each said pairing, the development means is arranged to perform a word order analysis and to reject future pairings which would violate a word order criterion.

12. A method of obtaining new translation units for a computer translation system, from examples of source and corresponding target language texts, comprising:

analysing the texts to obtain dependency relationships between language units thereof;

matching words of one text against all those of the other, to generate scores;

pairing words of the respective texts using said scores; and

providing new translation units using the paired words, and language units in each of the languages derived from the analyses.

13. A computer natural language translation system, comprising:

means for inputting source language text;

means for outputting target language text;

transfer means for generating said target language text from said source language text using stored translation data generated from examples of source and corresponding target language texts,

characterised in that said stored translation data comprises a plurality of translation components, each comprising:

surface data representative of the order of occurrence of language units in said component;

dependency data related to the semantic relationship between language units in said component; and

the dependency data of language components of said source language being aligned with corresponding dependency data of language components of said target language,

and in that said transfer means is arranged to use said surface data of said source language in analysing the source language text, and said surface data of said target language in generating said target language text, and said dependency data in transforming the analysis of said source text into an analysis for said target language.

14. A computer language translation development system, for developing data for use in translation, comprising:

means for allowing corresponding source and target example texts to be linked into source and target language dependency graphs;

means for allowing corresponding translatable nodes of said source and target language dependency graphs representing translatable parts of the source and target texts to be aligned; and

means for automatically generating, from said source and target language dependency graphs, respective associated surface representative graph having a tree structure.

15. A computer program comprising code to execute on a computer to cause said computer to act as the system of any preceding claim.

16. Apparatus for inferring new translation units which will allow a given source text to translate as a given target text comprising,

a database of translation units;

means arranged to analyse both the source text and the target text into one or more alternative representations using these units;

means arranged to indicate and score lexical alignments between the source and target texts;

means arranged to select one of the alternative source analyses and one of the alternative target analyses based on the scored alignments; and

means arranged to infer one or more translation units based on the selected source analysis, the target analysis and the alignment.

17. Apparatus according to claim 16 wherein said alternative representations are tree representations or representations that can be converted into tree representations.